Hotel Bookings: EDA with Data Visualization

Dataset Background & Structure

This dataset is reservation information of two hotels in Portugal. The arrival date of reservations is between July 1st, 2015 and August 31st, 2017. It provides 32 attributes and 119,390 reservations. The details in regards to each attributes can be found here

Quality Issues

Some quality issues that we can see on a quick analysis are:
- A number of missing data

Questions to Answer for Exploratory Data Analysis (EDA)

By analysing data to answer the following questions, we will have better understanding about our dataset.
- Where do the guests come from?
- How much do guests pay for a room per night?
- How does the price per night vary over the year?
- What are the busiest months?
- How long do people stay at the hotels?
- Bookings by market segment?
- How many bookings were cancelled?
- Which month have the highest number of cancellation?

Exploratory Data Analysis Goals

The dataset consists of 119,390 reservations. In order to discover some insights valuable for business decision, we will prepare dataset by cleaning missing values first, and move on to EDA process.

In EDA (Exploratory Data Analysis), we will compare the characteristics and transition of City Hotel with those of Resort Hotel as well as analyse overall trend of both hotels. For this,

we will utilize data visualization libraries such as Matplotlib, Seaborn, and Plotly generating box plot, choropleth map, line plot, bar plot, and pie chart.

Once the analysis process is complete, we can feed the important features extracted from the analysis with a machine learning model to improve the model's prediction performance.